CLASSIFICATION OF PHYTOCENOLOGICAL SAMPLES BY THE AID OF A COMPUTER

by

T. SIMON

Department of Plant Taxonomy and Ecology of the Eötvös Loránd University, Budapest

and

M. G. JURASSA

Bureau for Computation and for Mechanization of Building Administration, Budapest

Received on October 15th, 1969

The essential units of our native plant communities are usually evaluated by the long established methods of the so called Middle-Europian School (Zürich – Montpellier School) (see in Soó 1960, 1964, 1968). The use of a standard method (Braun – Blanquet 1928, 1951, Soó 1945) made it possible to compare the vegetation of the main part of Europe and of its countries.

Because of the nature of the structure of the vegetation and partly in consequence of the subjective use of the method, however, in the case of a great number of plantecological groups the results did not seem to be sufficient, not even on the basis of the traditional method. In certain cases there are many communities described wantonly, and in other cases the systematic division is not within the ranges of the "natural units".

Therefore the use of mathematical – statistical methods seemed to be warranted. All these methods (Gleason 1920, Czekanowski – Kulczynski 1927, Sörensen 1948, Goodall 1954, Poore 1955, Ellenberg 1956, Falinski 1960, Précsényi 1962, Ramsay 1964, Juhász-Nagy 1964, Simon 1965, Pócs 1966, etc. primary and general works) greatly contributed to the classification and the comparison of the units of the classification.

In the present paper we want to demonstrate an automatic method with the adequate adoption of the χ^2-test suitable for counting the contingency coefficient (Greig – Smith 1957, 1967, Juhász-Nagy 1964, Weber 1964). This method applies a computer (I.C.T. 1905) and a dendrical plotting of the results (Falinski 1960). Our method makes it possible to sort out a larger amount of samples (1 – 100 samples) obtained by the use of the long established methods, to select the similar samples and to establish a cenological estimation which is the closest approximation of the actual units.

Fig. 1. Rock sward (Minuartio – Festucetum pseudodalmaticae) and closed rock sward (Potentillo – Festucetum pseudodalmaticae) stands on the hill "Füzéri – Várhegy" (Photo: T. Simon)

Method

The material used was taken from the silicate rock sward communities of the Zemplén-mounties. From a great number of samples we chose the following ones:

4 samples from a rock sward stand *(Minuartio-Festucetum pseudo-dalmaticae)*

4 samples from a closed rock sward stand *(Potentillo-Festucetum pseudo-dalmaticae)*

2 samples from two different closed rock slope stands *(Genisto-Festucetum ovinae* nom. prov.). To make our example clear we took care to choose the rock sward communities and close rock sward communities to be different.

The data of the permanent quadrats based on a physiognomical basis can be found in Table I, where the samples are ordered on the basis of the selection of the differential species. In surveying the samples we examined the lichens, the mosses and the phanerogamous plants. The surface of one sample was 4m^2.

Mathematical – statistical principles

The aim of the present paper is to estimate the connection between single samples on the basis of the presence of the examined species. For this purpose we made a paired comparison of each samples and this process resulted in the following paired samples: $1-2$, $1-3$, $1-4$, ..., $1-10$; $2-3$, $2-4$, ..., $2-10$; ... ..., $9-10$. If the number of the samples in question is n, then that of the paired samples will be $p = \dfrac{n(n-1)}{2}$.

As a basis for the calculations we used the χ^2-test which is suitable for counting the contingency coefficient (the Yates-correction was not taken into consideration). The χ^2-test used is the following:

$$\chi^2 = \frac{(ad-bc)^2 N}{(a+b)(a+c)(b+d)(c+d)}$$

$$(\text{assumption: } a+b+c+d=N)$$

a is the number of the common species figuring in both samples in question
b is the number of the species which are present only in the 2^{nd} sample
c is the number of the species which are present only in the 1^{st} sample
d is the number of the species which are present neither in the 1^{st} nor in the 2^{nd} sample, but which occur in the rest
N is the number of all species studied.

In this case the presence or the absence of the species in the samples was regarded as the basis for the trend of the phenomenon (+ or –), but we did not take into consideration the amount of the species found in the samples.

In the calculations we had first constituted the paired samples and then we counted for each one the value of the χ^2-test and these results were implied in a table (χ^2 matrix).

Table 1.

Table of cenological samples ordered on the basis of differential species (1 – 4 = "*Minuartio-Festucetum pseudodalmaticae*", 5 – 8 = "*Potentillo-Festucetum pseudodalmaticae*", 9 – 10 = "*Genisto-Festucetum ovinae.*" nom. prov.)

Numbers of relevés	1	2	3	4	5	6	7	8	9	10
Localities of relevés	Füzér	Füzér	Füzér	Füzér	Füzér	Pusztafalu	Kovácsvágás	Füzér	Telkibánya	Telkibánya
Exposure	S	S	NE	NE	S	S	NE	S	SW	SW
Inclination	60	40	50	70	0	30	45	0	80	70
Cover of herbs layer	30	60	20	20	80	80	70	100	20	20
Cover of mossis layer	30	0	20	20	0	0	0	0	10	10
Height of vegetation in. cm.	30	30	20	30	40	40	30	40	30	30
Size of quadrats in sq. m.	4	4	4	4	4	4	4	4	4	4
Numbers of species of relevés	17	21	39	42	26	22	19	23	21	24
The species of the first "Association":										
Minuartia frutescens	2 – 3	3 – 4	1	+ – 1						
Hypnum cupressiforme	+	+	1	+ – 1						
Sempervivum hirtum	+ – 1	+	+	+ – 1						
S. marmoreum	+ – 1	+	+	+						
Parmelia conspersa	+	+	+	+						
Diploschistes scruposus	+	+	+	+						
Thymus marschallianus		+	+	+						
Saxifraga paniculata			1 – 2	2						
Camptothecium sericeum			+ – 1	+ – 1						
Medicago prostrata			+ – 1	+ – 1						
Peltigera rufescens			+ – 1	+ – 1						
Alyssum saxatile			+ – 1	+						
Hedwigia albicans			+ – 1	+						
Iris pumila			+ – 1	+						
Acarospora sp.			+	+						
Androsace elongata			+	+						
Cladonia furcata v. palamaea			+	+						
C. rangiformis			+	+						
Dicranum scoparium			+	+						
Draba verna			+	+						
Myosotis micrantha			+	+						
Parmelia sorediata			+	+						
Pertusaria sp.			+	+						
Rhizocarpon geographicum			+	+						
Rhytidium rugosum			+	+						
Saxifraga tridactylites			+	+						
Thuidium abietinum			+	+						

Table I.

Numbers of relevés	1	2	3	4	5	6	7	8	9	10
Localities of relevés	Füzér	Füzér	Füzér	Füzér	Füzér	Pusztafalu	Kovácsvágás	Füzér	Telkibánya	Telkibánya
The species of the second "Association":										
Trifolium alpestre					+	2	+	2 – 3		
Thymus pulegioides					+	+ – 1	1	1 – 2		
Teucrium chamaedrys					+	+ – 1	+	+ – 1		
Lactuca perennis					+	1	+			
Verbascum austriacum						+ – 1	+ – 1	+		
Achillea pannonica					+			+ – 1		
Asperula glauca					+	+ – 1				
Campanula divergentiformis					+	+ – 1				
Verbascum lychnitis					+ – 1			1		
Cynanchum vincetoxicum						+		+		
Inula hirta						+	+			
Stachys recta						+	+			
Fragaria vesca								1		
Phleum phleoides								1		
Poa angustifolia								1		
Genista elata								+ – 1		
Hypericum perforatum								+ – 1		
Potentilla argentea							+ – 1			
Primula veris							+ – 1			
Veronica austriaca								+ – 1		
The species of the third "Association":										
Festuca ovina									2	2
Cladonia squamata									1	+
Genista pilosa									+	+ – 1
Bartramia pomiformis									+	+
Rhacomitrium canescens									+	+
Solidago virga-aurea									+	+
Woodsia ilvensis									+	+
Galium vernum									+ – 1	
Carex digitata										+
Cephaloziella starkei									+	
Dicranum montanum									+	
Entodon schreberi										+
Hieracium pilosella										+
Hieracium umbellatum										+
Sieglingia decumbens										+

Table I.

Numbers of relevés	1	2	3	4	5	6	7	8	9	10
Localities of relevés	Füzér	Füzér	Füzér	Füzér	Füzér	Pusztafalu	Kovácsvágás	Füzér	Telkibánya	Telkibánya
Accompanying species:										
Festuca pseudodalmatica	+ – 1	+	1 – 2	1 – 2	4 – 5	4	5	3		
Euphorbia cyparissias		+	+ – 1	+	+ – 1	+	+ – 1	+		
Asplenium septentrionale	1	+	+	+		+			+	+
Seseli osseum	+	1	+ – 1	+	+ – 1	+ – 1				
Veronica orchidea		+	+	+				+ – 1	+ – 1	+
Allium montanum	+	1	1	+ – 1	1 – 2					
Potentilla arenaria	+	1 – 2	+ – 1	+	1					
Polytrichum piliferum	+	+				+			1 – 2	+
Ceratodon purpureus			+	+	+	+	+			
Poa pannonica ssp. scabra			+	+	1	+ – 1				
Sedum acre var. krajinae			+ – 1	1	+		+			
Polypodium vulgare			+ – 1	+					+	+
Asplenium trichomanes			+	+					+	+
Sedum maximum					+		+		+	+
Syntrichia ruralis			+ – 1	2			+			
Koeleria cristata			+ – 1	+ – 1				1		
Arenaria serpyllifolia			+	+			+			
Cardaminopsis arenosa				+					+	+
Cladonia chlorophaea			+	+						+
Rumex acetosella	+					+				+
Viscaria vulgaris								+ – 1		+
Dianthus carthusianorum				+		+				

Accidental species:

Alyssum alyssoides (7), Asperula cynanchica (7), Brachythecium rutabulum (5), Calamintha acinos (4), Campanula glomerata (8), Camptothecium lutescens (5), Catharinaea undulata (5), Centaurea axillaris (5), Cladonia pyxidata (9), Coronilla varia (6), Cotoneaster integerrima (2), Dicranella heteromalla (5), Euphorbia polychroma (7), Eurhynchium strigosum (5), E. zetterstädtii (5), Fissidens taxifolius (5), Linaria angustissima (6), Origanum vulgare (8), Parmelia caperata (9), Peucedanum cervaria (7), Poa nemoralis (9), Polytrichum juniperinum (6), Potentilla ascendens (5), Rosa canina (2), Torilis anthriscus (8).

It is well known that with the use of the χ^2-test it is possible to count P% values or contingency coefficient values (r). The equations used for these calculations are the following:

a)

$$P = \frac{\chi^2}{FG}$$

b) in the case of relation control with two classes qualitative variables

$$r = \sqrt{\frac{\chi^2}{N}}$$

c) in the case of relation control with more than two classes qualitative variables

$$r = \sqrt{\frac{\chi^2}{N(q-1)}}$$

(q is the value of the class number of the less value).

In the following we shall discuss the contingency coefficient on the basis of the following considerations. It is known that the contingency coefficient indicates a connection or similarity, and the value of r is within the ranges 0 and 1 ($0 \leq r \leq 1$). In the case of $0 < r < 0.4$ the connection between the two examined samples (or phenomenons) is slack; if $0.4 \leq r < 0.7$ then the connection is moderate; position $0.7 \leq r < 0.9$ indicates close connection and when the figure is $0.9 \leq r < 1$ the connection is very close (in the case of $r = 1$ the two phenomena compaired are identical; but $r = 0$ means an independence of phenomena).

In our examination of similarity, however, we did not calculate the values of contingency coefficient. From the connections related to the r values it is clear that r is proportional to the value of χ^2-test; if the value of χ^2-test is higher, that of r will be also higher. In consequence of the ratio mentioned above the higher value of χ^2-test will refer to a closer connection. In drawing the dendrit we took into consideration this fact.

Drawing of the dendrit

By drawing the dendrit we intended to point out the similarity or disparity (independence) of the samples, and the possibility of forming continguous groups. As mentioned above the higher value of the χ^2-test refers to a closer connection and the less value refers to a slack one or the lacking of the connection. In the χ^2 matrix we selected the highest values of the χ^2-test in the case of each paired sample, first by rows, then by columns. These two maximal values of single paired samples will point to the two other samples which are in the closest connection with the sample in question. Then from these paired values of the χ^2-test we selected the higher value. This served to establish the specific composition of the sample which is the most similar to that of the sample in question (primary connection). Any other value regarding this sample indicates a secondary connection.

Then we ordered the values of the χ^2-test regarding the primary and secondary connections according to their order of magnitude. This procedure gave three results:

a) we could estimate which sample is in the closest connection with one of the samples in question

b) among the samples which two show the greatest similarity (this paired sample will be the core of the dendrit) and finally

c) we could estimate the sequence of similarities.

Table II.

Data of the paired samples (*a* = the number of the common species; *b* = the number of the species which are present only in the 2^{nd} sample; *c* = the number of the species figuring only in the 1^{st} sample; *d* = the number of the species which are present neither in the 1^{st} sample nor in the 2^{nd} one, but which occur in the rest)

	a	b	c	d
1 — 2	16	1	5	87
1 — 3	11	6	28	64
1 — 4	11	6	31	61
1 — 5	4	13	22	70
1 — 6	5	12	17	75
1 — 7	1	16	18	74
1 — 8	1	16	22	70
1 — 9	3	14	18	74
1 — 10	4	13	18	74
2 — 3	14	7	25	63
2 — 4	14	7	28	60
2 — 5	5	16	21	67
2 — 6	5	16	16	72
2 — 7	2	19	17	71
2 — 8	3	18	20	68
2 — 9	4	17	17	71
2 — 10	4	17	18	70
3 — 4	39	0	3	67
3 — 5	8	31	18	52
3 — 6	6	33	16	54
3 — 7	6	33	13	57
3 — 8	4	34	18	53
3 — 9	5	34	14	56
3 — 10	5	34	17	53
4 — 5	8	34	18	49
4 — 6	6	35	15	53
4 — 7	6	36	13	54
4 — 8	4	37	18	50
4 — 9	6	36	19	48
4 — 10	6	36	16	51
5 — 6	11	15	11	72
5 — 7	9	17	9	74
5 — 8	7	19	16	67
5 — 9	1	25	20	63
5 — 10	1	25	21	62
6 — 7	10	12	9	78
6 — 8	6	15	16	72
6 — 9	2	19	19	69
6 — 10	3	19	18	69
7 — 8	6	13	17	73
7 — 9	1	18	20	70
7 — 10	1	18	21	69
8 — 9	1	22	20	66
8 — 10	2	21	20	66
9 — 10	15	6	7	81

In the case of the drawing of the dendrit we took into consideration the invert values of χ^2-test $\left(\frac{1}{\chi^2}\right)$ according to their order of magnitude. We connected the signs of the samples in question with an equidistant which was proportional to the invert value in accordance. On the dendrit the close (primary) connections were marked with continuus heavy lines and the slack (secondary) connections were marked with dotted lines. Obviously the lenght of the lines connecting the signs of the samples is proportional to the similarity of the samples. Taking into consideration the order of magnitude of the values we drew the dendrit which shows the following figure.

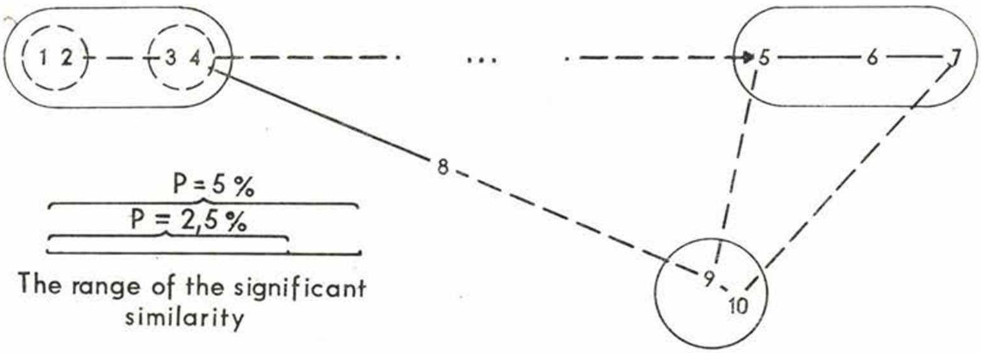

Fig. 2. The dendrit of the samples

In most cases it is possible to evaluate the connections correctly. In some critical cases it is advisible to compare the values of the χ^2-test counted with the values of an χ^2-table (it is advisable to take the level of $P = 5\%$, but rather of $P = 2.5\%$).

Use of the computer in the drawing and the computational process.

Taking into consideration the facts mentioned above one can state that a great amount of data and of calculation are needed in the process. In the present paper we had to calculate 45 values of the χ^2-test and 18 values of maximum. For example in the case of an investigation dealing with 50 samples one has to calculate 1225 values of the χ^2-test and 98 values of maximum (a person working 8 hours a day can realize this work with the aid of a desk-computer (HUNOR 131) about within 20 days).

It seemed to be practical to use a computer in the modelling process because of the reasons mentioned above. The program was written in FORTRAN algorithmic language oriented for I.C.T. 1905 computer. By the aid of the computer we can compare $1-100$ samples within a short-term (the calculations for the investigation described in the present paper required 4 minutes).

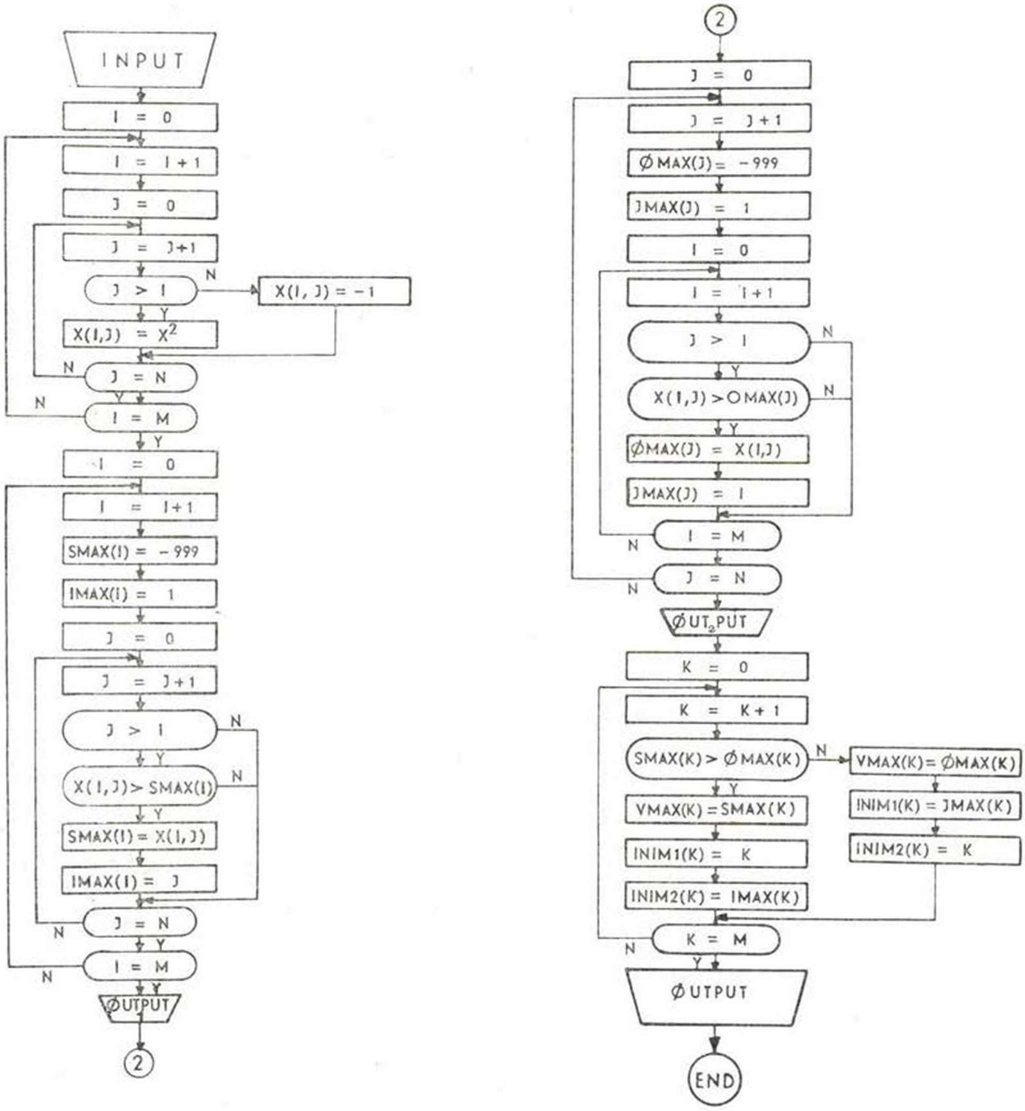

Fig. 3. The blockdiagram of the modell

The construction of the program is the following:

1. INPUT – by the aid of a subroutine it reads the data and the parameters from punch cards [W, A(I, J), B(I, J), C(I, J), D(I, J)].

Table III.

The highest and selected χ^2-values (data calculated by the computer)

OUTPUT 1.										
IMAX(I)	2	3	4	8	6	7	10	9	10	
SMAX(I)	72.55	10.80	96.88	4.44	10.37	15.04	3.17	4.17	42.40	
OUTPUT 2.										
JMAX(J)	1	2	3	4	5	6	4	5	9	
OMAX(J)	72.55	10.80	96.88	0.87	10.37	15.04	4.44	5.22	42.40	
OUTPUT										
INDM1(K)	1	1	3	3	5	6	6	4	9	9
INDM2(K)	2	2	4	4	6	7	7	8	10	10
VMAX(K)	72.55	72.55	96.88	96.88	10.37	15.04	15.04	4.44	42.40	42.40

2. MASTER segment

a) the values of the χ^2-test will be calculated and put into a matrix (χ^2 matrix; X(I, J)]

b) it selects the maximal χ^2-test values by rows (that is by paired samples) and puts them and their row and column index into vectors [OUTPUT 1; SMAX(I), IMAX(I)]

c) it selects the maximal χ^2-test values by columns and puts them and their row and column index into vectors [OUTPUT 2; OMAX(J), JMAX(J)]

d) it selects from the paired values of the χ^2-test the higher ones and puts them and their row and column index into vectors [VMAX(K), INDM1(K), INDM2(K)].

3. OUTPUT – the values and their index will be printed out on the line-printer.

Results

Examining Table I on the basis of the traditional method, according to the expectations of the differential species, the 1–4, 5–8 and 9–10 samples are segregatable (on the basis of constant and character species further estimation was not possible, because the cenological aspect of the communities has not been worked out). Within the group of 1–4 samples the 3 and 4 ones show significant difference and within the 5–8 samples the 5 one shows a moderate difference.

The classification of the samples supported by the calculation verifies the facts mentioned above only in broad outline. On the other hand the similarity of the 3–4 samples and their difference from the 1–2 samples are conspicuous and so on the basis of the slack connection calculated (dotted line on the dendrit) one can evaluate them as a separate association.

From the group of 5—8 samples which are homogeneous physiognomically the difference of the 8 sample is obvious. It has a closer connection only with the rock sward communities. It seems to be practical to omit it from the tabular drawing of the community as a transitional sample. The drawing clearly shows the close connection of the 9—10 samples and their difference from the other ones.

A further aim of our studies was to verify the practicability of our dendrical method. But the classification and cenological estimation of the given silicate rock sward communities or that of other ones will be possible only after determining their constant and character species and after classifying a considerably larger amount of their sample material by the aid of the method described above.

Summary

In the present paper we classified samples obtained by the aid of the long established permanent quadrat method. In the process we made a paired comparison of each sample on the basis of the common (a), of the own (b, c) and of the remaining (d) species of the paired samples by using the χ^2-test suitable for the calculation of the contingency coefficient and by the aid of a computer (I.C.T. 1905). The use of the computer speeded up the method considerably. With the dendrical drawing (or in critical cases with the aid of a Table of χ^2-test values) of the invert results it is possible to control and to revise the traditional method of cenological classification. The method is suitable for a speedy, unified and objective classification of materials of "continuus" samples.

REFERENCES

Braun-Blanquet, J. 1928, 1951. Pflanzensoziologie (Plantecology), Wien.

Czekanowski, J. 1909. Zur Differencialdiagnose der Neanderthalgruppe (Notes to the differentialdiagnosis of the Neanderthal groups). Korrespondenzblatt d. D. G. f. Anthr. un Urgesch. **11:** (6/7).

Ellenberg, H. 1956. Aufgaben und Methoden der Vegetationskunde (Problems and Methods of the Vegetation Research). In Walter, H.: Einführung in die Phytologie (Introduction to the Phytology) IV/1. Stuttgart

Falinski, J. 1960. Zastosowanie taksonomii wroclawskiej do fitosocjologii (The use of Wroclaw School's taxonomy in the plantecology). Acta. Soc. Bot. Pol. **29:** 333—361.

Gleason, H. A. 1920. Some applications of the quadrat method. Bull. Torrey Bot. Cl. **47:** 21—43.

Gooddal, D. W. 1953. Objective methods for the classification of vegetation. I. The use of positive interspecific correlation. Austr. J. Bot. **1:** 39—63.

Gooddal, D. W. 1954. Vegetational classification and vegetational continua. Angew. Pflanzsoz. **1:** 168—182.

Goodman, L. A. — Kruskal, W. H. 1959. Measures of Association for Cross Classifications. Journ. of the Am. Stat. Ass. Dec. 1954. **49:** 732—764.

Goodman, L. A. — Kruskal, W. H. 1959. Measures of Association for Cross Classifications. II.: Further Discussion and References. Journ. of the Am. Stat. Ass. March. 1959. **54:** 123—163.

Greig-Smith, M. A. 1957, 1964. Quantitative Plant Ecology. London.

Juhász-Nagy, P. 1964. Investigations on the Bulgarian Vegetation. Some hygrophilous Plant Communities (I–III). Acta Biol. Debrecina 2: 47–70.

Kulczynski, S. 1927. Zespoly roslin w Pieninach (The plant communities of the Pienins). Bull. Int. Acad. Pol. Sc. Lett. Sér. B. Suppl. 2: 57–203.

Poore, M. E. D. 1955. The use of phytosociological methods in ecological investigation. I. . The Braun-Blanquet system; II. Practical issues involved in an attempt to apply the Braun-Blanquet system. Journ. of Ecol. 43: 226–244; 245–269.

Pócs, T. 1966. Statisztikus matematikai módszer növénytársulások elhatárolására (Statistical-mathematical method for denotation of plant communities). Acta Acad. Paed. Agr. Nova Ser. 4: 441–454.

Précsényi, I. 1962. Kvantitatív cönológiai vizsgálatok *Festucetum vaginatae*-ban. (Quantitative cenological investigations in *Festucetum vaginatae)* Doktori értekezés (Dissertation).

Ramsay, D. Mc. C. 1964. An analysis of Nigerian savanna II. An alternative method of analysis and its application to Gombe sandstone vegetation. Journ. of Ecol. 52: 457–466.

Simon, T. 1965. Über die *Seslerietum rigidae*-Assoziationen in Siebenbürgen (The *seslerietum rigidae*-communities in Transylvania). Acta Bot. Acad. Sci. Hung. 11: 221–234.

Soó, R. 1945. Növényföldrajz (Plant Ecology). Budapest.

Soó, R. 1960. Bibliographia Phytosociologica: Hungaria. Excerpta Botanica Sect. B. 2: 93–156.

Soó, R. 1964, 1968. A magyar flóra és vegetáció rendszertani-növényföldrajzi kézikönyve (Plant taxonomical and ecological manual of the Hungarian flore and vegetation). I., III. Budapest, Akadémiai Kiadó.

Sörensen, T. 1948. A method groups of esteblishing of equal amplitudo in plant sociology based on similarity of species content. Kong. Dans. Vidensk. Selskab. Biol. Skr. 5: 1–34.

Sváb, J. 1967. Biometriai módszerek a mezőgazdasági kutatásban (Biometrical methods in the agricultural research). Budapest.

Weber, E. 1964. Grundriss der biologischen Statistik (Manual of biological statistic). Jena·